

# Global Effects of Feature-based Attention Depend on Surprise

*(Abbreviated title: Feature-based Attention Depends on Surprise)*

Cooper A. Smout<sup>\*1,2</sup>, Marta I. Garrido<sup>1,2,3,4</sup>, and

Jason B. Mattingley<sup>1,2,5,6</sup>

<sup>1</sup>Queensland Brain Institute, University of Queensland, Brisbane, Australia, 4072

<sup>2</sup>Australian Research Council Centre of Excellence for Integrative Brain Function, Australia, 3800

<sup>3</sup>Melbourne School of Psychological Sciences, University of Melbourne, Melbourne, Australia, 3010

<sup>4</sup>Centre for Advanced Imaging, University of Queensland, Brisbane, Australia, 4072

<sup>5</sup>School of Psychology, University of Queensland, Brisbane, Australia, 4072

<sup>6</sup>Canadian Institute for Advanced Research (CIFAR), Toronto, Canada, M5G1M1

\* cooper.smout@gmail.com

Number of pages: 34

Number of figures: 7

Abstract: 233 words

Introduction: 602 words

Discussion: 1498 words

This research was supported by the Australian Research Council (ARC) Centre of Excellence for Integrative Brain Function (ARC Centre Grant CE140100007). J.B.M was supported by an ARC Australian Laureate Fellowship (FL110100103). M.I.G. was supported by a University of Queensland Fellowship (2016000071). The authors declare no competing financial interests.

24 **Abstract**

25 Recent studies have shown that prediction and attention can interact under various  
26 circumstances, suggesting that the two processes are based on interdependent neural  
27 mechanisms. In the visual modality, attention can be deployed to the location of a task-relevant  
28 stimulus ('spatial attention') or to a specific feature of the stimulus, such as colour or shape,  
29 irrespective of its location ('feature-based attention'). Here we asked whether predictive  
30 processes are influenced by feature-based attention outside the current spatial focus of  
31 attention. Across two experiments, we recorded neural activity with electroencephalography  
32 (EEG) as human observers performed a feature-based attention task at fixation and ignored a  
33 stream of peripheral stimuli with predictable or surprising features. Central targets were  
34 defined by a single feature (colour or orientation) and differed in salience across the two  
35 experiments. Task-irrelevant peripheral patterns usually comprised one particular conjunction  
36 of features (standards), but occasionally deviated in one or both features (deviants). Consistent  
37 with previous studies, we found reliable effects of feature-based attention and prediction on  
38 neural responses to task-irrelevant patterns in both experiments. Crucially, we observed an  
39 interaction between prediction and feature-based attention in both experiments: the neural  
40 effect of feature-based attention was larger for surprising patterns than it was for predicted  
41 patterns. These findings suggest that global effects of feature-based attention depend on  
42 surprise, and are consistent with the idea that attention optimises the precision of predictions  
43 by modulating the gain of prediction errors.

44 **Significance Statement**

45 Two principal mechanisms facilitate the efficient processing of sensory information: *prediction*  
46 uses prior information to guide the interpretation of sensory events, whereas *attention* biases

47 the processing of these events according to their behavioural relevance. A recent theory  
48 proposes to reconcile attention and prediction under a unifying framework, casting attention as  
49 a ‘precision optimisation’ mechanism that enhances the gain of prediction errors. Crucially,  
50 this theory suggests that attention and prediction interact to modulate neural responses, but this  
51 hypothesis remains to be tested with respect to feature-based attention mechanisms outside the  
52 spatial focus of attention. Here we show that global effects of feature-based attention are  
53 enhanced when stimuli possess surprising features, suggesting that feature-based attention and  
54 prediction are interdependent neural mechanisms.

## 55 **Introduction**

56 Selective attention mechanisms enhance the processing of sensory stimuli that are relevant for  
57 guiding behaviour (Desimone & Duncan, 1995; Posner, 1994). Visual processing can be biased  
58 toward stimuli at a relevant location, commonly known as ‘spatial attention’, or toward stimuli  
59 that possess task-relevant features, known as ‘feature-based attention’ (Carrasco, 2011).  
60 Monkey neurophysiology studies (Martinez-Trujillo & Treue, 2004) and human neuroimaging  
61 studies (Gledhill, Grimsen, Fahle, & Wegener, 2015; Saenz, Buracas, & Boynton, 2002;  
62 Serences & Boynton, 2007) have demonstrated that neural responses to stimuli at task-  
63 irrelevant locations are enhanced when they possess task-relevant features, demonstrating that  
64 the effects of feature-based attention are global and dissociable from those of spatial attention.  
65 In humans, neural responses to visual stimuli at task-irrelevant locations can be enhanced when  
66 they possess surprising features (e.g., colour, orientation, motion), demonstrating that top-  
67 down ‘prediction’ mechanisms also exert a global effect on incoming sensory signals (Friston,  
68 2005; Stefanics, Kremláček, & Czigler, 2014). At present, it is unknown whether the global  
69 effects of visual feature-based attention can interact with those of prediction. Here we used

70 electroencephalography (EEG) to measure neural responses to peripheral visual stimuli that  
71 were predictable or surprising along two feature dimensions (orientation and colour), and tested  
72 whether attending to a particular feature at fixation modulated the effect of prediction on neural  
73 responses to peripheral stimuli at task-irrelevant locations.

74 Predictive coding theories propose that top-down prediction signals effectively ‘silence’  
75 bottom-up sensory signals that match the predicted content, leaving only the remaining  
76 *prediction error* to propagate forward and update a model of the sensory environment (Friston,  
77 2005; Rao & Ballard, 1999). In addition to predicting the *content* of sensory signals, an optimal  
78 inference system should also estimate the level of uncertainty about its predictions (i.e., inverse  
79 precision; Hohwy, 2012). Recently, it has been proposed that selective attention mechanisms  
80 fulfil this role, optimising the expected precision of predictions by enhancing the activity of  
81 units encoding prediction errors for attended stimuli (Feldman & Friston, 2010; Friston, 2009,  
82 2010). Recent studies have supported this theory by demonstrating that selective attention and  
83 prediction can interact under various circumstances (Auksztulewicz & Friston, 2015; Jiang,  
84 Summerfield, & Egnér, 2013; Kok, Rahnev, Jehee, Lau, & De Lange, 2012; Marzecová,  
85 Widmann, SanMiguel, Kotz, & Schröger, 2017; Smout, Tang, Garrido, & Mattingley, 2019).  
86 However, selective attention mechanisms encompass distinct information-processing  
87 subcomponents (e.g., spatial attention, temporal attention) across sensory modalities (e.g.,  
88 auditory, visual) and it is important to establish which of these subcomponents interacts with  
89 prediction and in what manner. In the visual domain, previous studies that reported an  
90 interaction between attention and prediction typically presented stimuli at task-relevant  
91 locations (Jiang et al., 2013; Kok, Rahnev, et al., 2012; Marzecová et al., 2017; Smout et al.,  
92 2019). One previous study found an effect of feature-based attention on mismatch responses to  
93 stimuli at task-irrelevant locations, but this study presented clearly visible targets that likely

94 did not necessitate a tight focus of spatial attention on the central stimulus stream (Czigler &  
95 Sulykos, 2010). Thus, it remains unclear whether prediction can interact with global feature-  
96 based attention mechanisms that modulate neural responses to stimuli outside the spatial focus  
97 of attention.

98 Here we tested whether feature-based attention modulates the effect of prediction at task-  
99 irrelevant locations by comparing event-related potentials evoked by peripheral stimuli that  
100 either matched ('congruent') or mismatched ('incongruent') a cued feature of the target at  
101 fixation. Participants searched for targets at fixation while predictable or surprising task-  
102 irrelevant stimuli were presented in the periphery. We conducted two experiments that differed  
103 in the salience of central targets and distractors to investigate whether the strength of the top-  
104 down feature-set modulates the neural interaction between feature-based attention and  
105 prediction.

## 106 **Methods**

### 107 **Experiment 1: Effects of feature-based attention at fixation on neural responses to** 108 **predicted and surprising peripheral stimuli**

#### 109 *Participants*

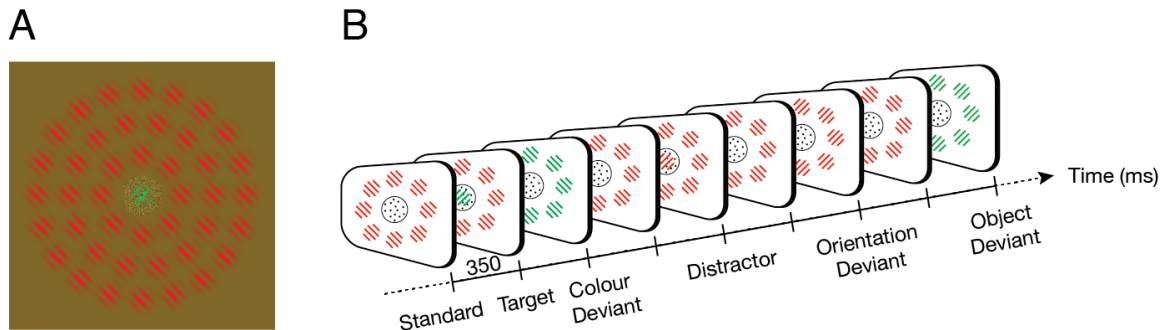
110 Twenty-four healthy adults (13 female, 11 male, age =  $22.08 \pm 2.38$  years) with normal or  
111 corrected-to-normal vision were recruited for participation via an online portal at The  
112 University of Queensland. The study was approved by The University of Queensland Human  
113 Research Ethics Committee, and all participants provided written, informed consent before  
114 commencing the experiment.

115 *Stimuli and apparatus*

116 Participants were positioned at a viewing distance of 57 cm and seated in a comfortable  
117 armchair in an electrically shielded laboratory. Stimuli were presented on a 61 cm LED monitor  
118 (Asus, VG248QE) with a 1920 x 1080 pixel resolution and refresh rate of 120 Hz, using  
119 PsychToolbox presentation software (Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray,  
120 R., & Broussard, 2007) for Matlab (v.15b) running under Windows 7 with an NVidia Quadro  
121 K4000 graphics card. The intensity of the green phosphor was adjusted per participant to  
122 produce subjective equiluminance with that of the red phosphor at full intensity. The  
123 equiluminance point was determined prior to the experiment using minimum motion  
124 photometry (Anstis & Cavanagh, 1983), with intensity values determined by two interleaved  
125 adaptive staircases (1 up-1 down, stopping after 15 reversals).

126 Central and peripheral stimuli were sinusoidal Gabors (diameter:  $4.72^\circ$ , spatial frequency:  $0.94$   
127  $c^\circ$ , 100% contrast) with one of two orientations (tilted  $45^\circ$  clockwise or counterclockwise of  
128 vertical) and one of two colours (red or green). Central stimuli were superimposed over a red-  
129 green noise patch (diameter:  $4.72^\circ$ ) and onset every 700 - 1400 ms for 66.67 ms. Twenty percent  
130 (20%) of central stimuli were targets (approximately 28 targets and 114 distractors per block).  
131 Multi-element peripheral stimuli ('patterns') were arranged in three concentric circles (radii:  
132  $4.72^\circ$ ,  $8.49^\circ$ ,  $12.26^\circ$ ; containing 8, 14, and 20 Gabors, respectively; *Figure 1A*). Peripheral  
133 patterns were presented every 350 ms for 66.67 ms (428 events per block) on top of a  
134 background that alternated between uniform red and green pixels at the screen refresh rate (120  
135 Hz), producing a subjective percept of a uniform brown background. During each block,  
136 peripheral patterns were more likely to contain one of the four possible feature conjunctions  
137 (e.g. clockwise-tilted red Gabors; 76% of presentations, *standards*), with the other three feature

138 conjunctions being rare and of equal likelihood (8% each, *deviants*). Standards were pseudo-  
139 randomized across blocks, and the order of deviants was pseudo-randomized within blocks.



140

141 *Figure 1. Stimulus display and task. (A)* Stimulus display, showing a green clockwise-tilted target  
142 within the central noise patch and a peripheral pattern consisting of three concentric rings of red  
143 counterclockwise-tilted Gabors. *(B)* Simplified task diagram. In this example trial, participants  
144 monitored for green Gabors (targets) and ignored red Gabors (distractors) within the central noise  
145 patch. Peripheral patterns typically contained red counterclockwise-tilted Gabors (standards), but  
146 occasionally contained red clockwise-tilted Gabors (orientation deviant), green counterclockwise-tilted  
147 Gabors (colour deviant), or green clockwise-tilted Gabors (object deviant, i.e. deviating in both colour  
148 and orientation). In this example trial, the colour and object deviants shared features with the target  
149 (i.e., green) and would thus be labelled 'congruent'.

## 150 *Procedure*

151 Participants were asked to fixate on a central dot and click a mouse button as soon as they  
152 detected a target in the stream of central stimuli, continuously throughout blocks (duration: 150  
153 s), while ignoring central distractors and peripheral patterns (*Figure 1B*). In each block, Gabor  
154 targets were designated as either (1) red, (2) green, (3) clockwise-tilted, or (4)  
155 counterclockwise-tilted. Note that each condition dictated two of the four possible feature  
156 conjunctions as targets and two as distractors (e.g., if searching for clockwise-tilted targets,  
157 both red and green clockwise-tilted Gabors were valid targets).

158 Participants completed two practice blocks with auditory feedback after each response, before  
159 being fitted with the EEG cap and electrodes (see *EEG Data Acquisition*). Participants then  
160 completed 16 test blocks with target type and standard pattern features pseudorandomized

161 across blocks (6848 peripheral patterns per session). Feedback on mean reaction time and the  
162 number of hits and false alarms was provided between blocks.

### 163 *Behavioural Data Analysis*

164 We investigated whether the feature-congruence and predictability of peripheral patterns  
165 affected participants' detection of central targets. Targets were sorted into *prediction*  
166 *conditions* according to whether the preceding pattern (i.e., the peripheral stimulus presented  
167 up to 700 ms prior to peak target contrast) was a standard ('predicted') or a deviant  
168 ('surprising'), and *feature-congruence conditions* according to whether the preceding  
169 peripheral pattern matched the features of the central target ('congruent') or distractor  
170 ('incongruent'). Participant responses were scored as hits if they occurred within 1 s of the  
171 onset of a target. Successive responses within this window were ignored, as were any responses  
172 that occurred within 250 ms of a preceding response. Because we observed differences in hit  
173 rates and reaction times between target feature conditions (i.e., the feature that participants  
174 searched for at fixation, e.g. 'red'), we first normalised hit rates and reaction times within each  
175 target feature condition, separately for feature-congruence and prediction conditions, and then  
176 collapsed across the target feature conditions. The resulting normalised hit rates and reaction  
177 times were then subjected to two-way repeated measures ANOVAs to assess the effects of  
178 peripheral pattern prediction (two levels: predicted, surprising) and feature-congruence (two  
179 levels: congruent, incongruent) on target detection.

### 180 *EEG Data Acquisition*

181 Participants were fitted with a 64 Ag-AgCl electrode EEG system (BioSemi Active Two:  
182 Amsterdam, Netherlands). Continuous data were recorded using BioSemi ActiView software  
183 (<http://www.biosemi.com>), and were digitized at a sample rate of 1024 Hz with 24-bit A/D



184 conversion and a .01 – 208 Hz amplifier band pass. All scalp electrode offsets were adjusted  
185 to below 20 $\mu$ V prior to beginning the recording. Pairs of flat Ag-AgCl electro-oculographic  
186 electrodes were placed on the outside of both eyes, and above and below the left eye, to record  
187 horizontal and vertical eye movements, respectively.

### 188 *EEG Preprocessing*

189 EEG recordings were processed offline using the EEGLab toolbox in Matlab (Delorme &  
190 Makeig, 2004). Data were resampled to 256 Hz and high-pass filtered with a passband edge at  
191 0.5 Hz (1691-point Hamming window, cut-off frequency: 0.25 Hz, -6 db). Raw data were  
192 inspected for the presence of faulty scalp electrodes (none were found). To clean the data, we  
193 applied an iterative process of artifactual epoch and component rejection using independent  
194 component analyses (ICA). The data were segmented into 350 ms epochs surrounding Gabor  
195 onsets (50 ms pre- and 300 ms post-stimulus) and baseline activity prior to stimulus onset was  
196 removed from each epoch. Epochs were subjected to ICA, and the SASICA plugin for EEGLab  
197 (Chaumon, Bishop, & Busch, 2015) was used to identify blink, saccade, and focal trial  
198 components. Epochs were rejected if they met any of the following criteria: (1) blink  
199 component activity greater than  $\pm 10$   $\mu$ V between -50 and 150ms; (2) saccade component  
200 activity greater  $\pm 5$   $\mu$ V between 0 and 350 ms; (3) focal component activity exceeding a joint  
201 probability threshold of  $\pm 7$  SD (5.5% of epochs were removed due to blink, saccade, or focal  
202 activity). The remaining epochs were then subjected to ICA for a second time, and SASICA  
203 was used again to identify artifactual components.

204 For further analysis, the resampled raw data were band-pass filtered between 0.5 and 40 Hz  
205 (1691-point Hamming window, cut-off frequencies: 0.25 and 40.25 Hz, -6 db) and segmented  
206 into 550 ms epochs surrounding Gabor onsets (100 ms pre- and 450 ms post-stimulus). Epochs

207 containing artefacts (identified previously using the first ICA) were removed. Independent  
208 component weights from the second ICA were applied to this new dataset and artefactual  
209 components (identified previously using the second ICA) were removed. Baseline activity in  
210 the 100 ms prior to stimulus onset was removed from each epoch.

### 211 *Event-Related Potential and Bayes Factor Analyses*

212 Peripheral patterns were sorted into prediction conditions based on whether they were  
213 standards (repeated at least 4 times; ‘predicted’) or deviants (‘surprising’), and attention  
214 conditions based on whether they shared features with central targets (‘congruent’) or  
215 distractors (‘incongruent’) in the central task. Trials in each attention and prediction condition  
216 were averaged within participants to produce event-related potentials (ERPs) for each  
217 individual. Statistical analyses of condition ERPs were conducted using two-tailed cluster-  
218 based permutation tests across participants (Monte-Carlo distribution with 5000 permutations,  
219  $p_{cluster} < 0.05$ ; sample statistic: dependent samples  $t$ -statistic, aggregated using the maximum  
220 sum of significant adjacent samples,  $p_{sample} < .05$ ) in the Fieldtrip toolbox for Matlab  
221 (Oostenveld, Fries, Maris, & Schoffelen, 2011). Statistical analyses of univariate condition  
222 averages were conducted using paired-samples  $t$ -tests and Bayesian analyses. The Bayes factor  
223 analyses allowed for quantification of evidence in favour of either the null or alternative  
224 hypothesis, with  $BF_{10} > 3$  indicating substantial support for the alternative hypothesis and  $BF_{10}$   
225  $< 0.33$  indicating substantial support for the null hypothesis. Bayes factors were computed  
226 using the Dienes (2014) calculator in Matlab.

227 **Experiment 2: Replication with individually thresholded manipulation of feature-based**  
228 **attention at fixation**

229 In Experiment 1, the high contrast targets were detected at near-ceiling levels. To investigate  
230 whether the neural interaction between feature-based attention and prediction is sensitive to the  
231 strength of the top-down feature set, we conducted a second experiment in which central targets  
232 and distractors were individually thresholded to be less salient. Except for the minor  
233 methodological differences noted below, Experiment 2 was the same as Experiment 1 and thus  
234 afforded an opportunity to replicate the original results in a separate group of participants.

235 *Methods*

236 A new cohort of 24 healthy adults with normal (or corrected-to-normal) vision was recruited  
237 to participate in Experiment 2 (12 female, 12 male, age =  $22.17 \pm 2.88$  years, mean  $\pm$  SEM).  
238 The stimuli and apparatus were identical to those used in Experiment 1 (*Figure 1*), except that  
239 in Experiment 2 the central stimuli (targets and distractors) were presented at lower contrast  
240 and with a sinusoidal onset and offset profile (total duration: 700 ms). The peak contrast of the  
241 central stimuli was determined during the two practice blocks, using a transformed and  
242 weighted up/down adaptive staircase configured to approximate 83% detection of targets  
243 (up/down step ratio: 1/3, up/down size ratio: .1/.07; García-Pérez, 1998). Blocks lasted for 150  
244 s (as per Experiment 1) for all except two participants, for whom blocks lasted for 120 s (due  
245 to time constraints for these two individuals). Participant responses were scored as hits if they  
246 occurred within 1 s of the peak target contrast (i.e., within 1.35 s of target onset, accounting  
247 for the 350 ms on-ramp). During EEG preprocessing, we interpolated 11 faulty electrodes  
248 (across 5 participants) using the average activation across neighbouring electrodes (defined by  
249 the EEGlab Biosemi 64 template) and removed 4.1% of epochs due to blink, saccade, or focal  
250 component contamination.

251

## Results

### 252 **Experiment 1: Effects of feature-based attention at fixation on neural responses to** 253 **predicted and surprising peripheral stimuli**

#### 254 *Feature-Congruent Peripheral Patterns Interfere with Target Detection at Fixation*

255 We first asked whether the congruence between peripheral pattern and central target features  
256 affected participants' detection of central targets shortly after pattern onset. There was no  
257 significant main effect of feature-congruence on normalized hit rates (congruent =  $94.41 \pm$   
258  $1.28\%$ ,  $0.06 \pm 0.19$  z-normalised, incongruent =  $93.63 \pm 1.39\%$ ,  $-0.08 \pm 0.20$  z-normalised,  
259  $F(1,23) = 3.33$ ,  $p = .081$ ,  $\eta_p^2 = .013$ ). There was a significant main effect of feature-congruence  
260 on normalised reaction times, however, with participants responding more slowly to central  
261 targets preceded by congruent peripheral patterns ( $438.80 \pm 9.43$  ms, mean  $\pm$  SEM;  $0.04 \pm 0.19$   
262 z-normalised) than to those preceded by incongruent peripheral patterns ( $436.09 \pm 9.53$  ms; -  
263  $0.04 \pm 0.20$  z-normalised,  $F(1,23) = 5.70$ ,  $p = .026$ ,  $\eta_p^2 = 0.20$ ). This finding suggests that  
264 participants were more distracted by peripheral patterns with task-relevant (congruent)  
265 features, relative to those with task-irrelevant (incongruent) features, and is consistent with the  
266 theory that involuntary orienting to task-irrelevant stimuli is contingent on attentional control  
267 settings (Folk, Remington, & Johnston, 1992).

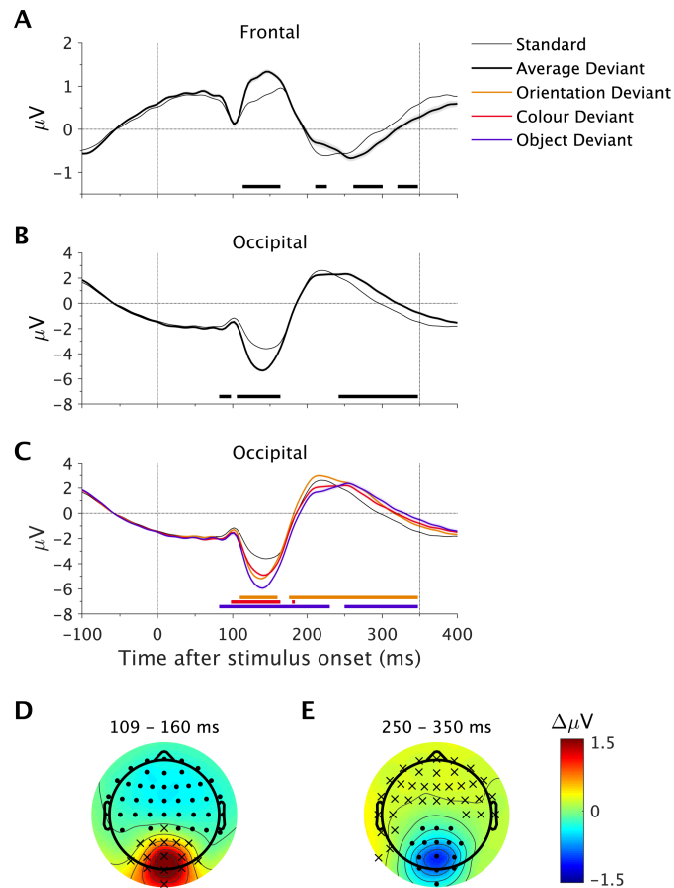
#### 268 *Peripheral Pattern Prediction Does Not Affect Target Detection*

269 In a second analysis we asked whether the predictability of peripheral patterns affected  
270 behavioural responses to subsequent central targets. There was no significant effect of  
271 peripheral pattern prediction on normalised hit rates (predicted =  $94.39 \pm 1.22\%$ ,  $0.04 \pm 0.18$ ,  
272 surprising =  $93.65 \pm 1.47\%$ ,  $-0.06 \pm 0.21$ ,  $F(1,23) = 1.63$ ,  $p = .215$ ,  $\eta_p^2 = .07$ ) or normalised  
273 reaction times (predicted =  $437.12 \pm 9.66$  ms,  $-0.01 \pm 0.20$ , surprising =  $437.76 \pm 9.32$  ms,  $0.01$

274  $\pm 0.19$ ,  $F(1,23) = 0.27$ ,  $p = .605$ ,  $\eta_p^2 = .01$ ), and no interaction between prediction and feature-  
275 congruence on either normalised hit rates ( $F(1,23) = 0.31$ ,  $p = .582$ ,  $\eta_p^2 = .01$ ) or normalised  
276 reaction times ( $F(1,23) = 0.15$ ,  $p = .701$ ,  $\eta_p^2 = .01$ ). These findings suggest that the  
277 predictability of peripheral patterns did not modulate the extent to which participants were  
278 distracted from their task at fixation.

### 279 *Prediction Decreases Neural Activity*

280 We next assessed the main effect of prediction on neural activity by comparing ERPs to  
281 peripheral deviant patterns ('surprising' patterns, collapsed across orientation, colour, and  
282 object deviants) and standard patterns that had been repeated at least 4 times ('predicted'  
283 patterns). Relative to baseline, standards evoked smaller neural responses than deviants (*Figure*  
284 2). Over posterior electrodes, the neural response to standards was significantly reduced  
285 relative to deviants during both the early negative deflection (i.e. standards > deviants; 82 - 164  
286 ms,  $p = .020$ ) and the late positive deflection (i.e. deviants > standards < deviants; 242 - 348  
287 ms,  $p = .010$ ; *Figure 2B*). Over frontal electrodes, the neural response to standards was  
288 significantly reduced relative to deviants during both the early positive deflection (i.e.  
289 standards < deviants; 90 - 230 ms,  $p < .001$ ) and the late negative deflection (i.e. standards >  
290 deviants; 254 - 348 ms,  $p = .008$ ; *Figure 2A*). These effects are consistent with the theory that  
291 surprising stimuli (deviants) produce greater prediction errors than predicted stimuli  
292 (standards; Friston, 2005, 2009; Rao & Ballard, 1999).



293

294 *Figure 2.* Main effect of prediction in Experiment 1. **(A-B)** ERPs evoked by standards and deviants  
295 (collapsed across deviant types) at frontal electrodes (Fz, F1, F3, AFz, AF3, AF4; **A**) and occipital  
296 electrodes (Oz, O1, O2, POz, PO3, PO4; **B**). Shading indicates the within-subject standard error of  
297 the mean, calculated relative to standards. Black bars along the x-axis denote significant timepoints at  
298 the displayed electrodes (cluster-corrected). **(C)** ERPs evoked by standards and each of the three  
299 deviant conditions. Shading indicates the within-subject standard error of the mean, calculated  
300 separately for each deviant condition relative to standards. Yellow, red and purple bars along the x-  
301 axis denote significant differences between standards and each corresponding deviant condition  
302 (cluster-corrected). **(D-E)** Headmaps show the effect of prediction (standard minus average deviant)  
303 during the indicated time windows. Asterisks and dots denote electrodes with larger or smaller  
304 responses, respectively, across at least 33% of the averaged time points (cluster-corrected).

305 We followed up this result with direct comparisons between standards and each type of deviant,  
306 which revealed similar effects to those reported above for the average deviant condition  
307 (*Figure 2C*). Early posterior negativities were smaller in response to standards than orientation  
308 deviants (109 – 160 ms,  $p = .033$ ), colour deviants (98 – 164 ms,  $p = .040$ ), and object deviants  
309 (82 – 348 ms,  $p < .001$ ), and late posterior positivities were significantly smaller in response to  
310 standards than orientation deviants (176 – 348 ms,  $p < .001$ ) and object deviants (250 – 348

311 ms,  $p = .019$ ). Early frontal positivities were smaller in response to standards than orientation  
312 deviants (98 – 164 ms,  $p = .002$ ), colour deviants (86 – 238 ms,  $p = .001$ ), and object deviants  
313 (102 – 238 ms,  $p < .001$ ), and late frontal negativities were smaller in response to standards  
314 than orientation deviants (242 – 348 ms,  $p < .001$ ) and object deviants (84 – 348 ms,  $p < .001$ ).

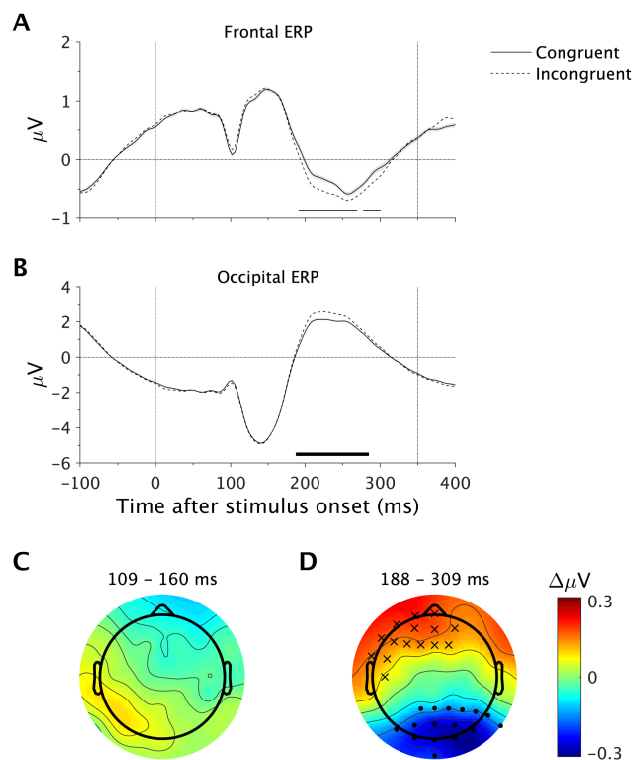
### 315 *Visual Mismatch Negativities Are Additive Across Feature Deviations*

316 Because previous investigations have suggested that the visual mismatch negativity (vMMN)  
317 is non-additive across feature deviations (Czigler & Sulykos, 2010), we also tested for  
318 differences between vMMNs evoked by each type of deviant (orientation, colour, or object).  
319 We used a data-driven approach to identify spatiotemporal samples (electrodes x timepoints)  
320 that were significantly different from standards in all three deviant conditions (electrodes: Pz,  
321 P1, P2, P3, P4, POz, PO3, PO4, Oz, O1, O2, Iz; timepoints: 109 – 160 ms) and then averaged  
322 across these samples to produce one amplitude value per deviant condition and participant. We  
323 then compared each pair of deviant conditions with paired-samples  $t$ -tests and Bayesian  
324 analyses, using a uniform prior with upper and lower bounds set to the average vMMN  
325 amplitude. As can be seen in *Figure 2C*, there was no difference between the orientation ( $-1.02$   
326  $\pm 0.13$   $\mu\text{V}$ ) and colour vMMN ( $-0.98 \pm 0.14$   $\mu\text{V}$ ,  $t(23) = -0.35$ ,  $p = .733$ ,  $BF_{10} = 0.14$ ). In  
327 contrast, the object vMMN ( $-1.64 \pm 0.16$   $\mu\text{V}$ ) was significantly larger than both the orientation  
328 vMMN ( $t(23) = -5.39$ ,  $p < .001$ ,  $BF_{10} = 2.4 \times 10^5$ ) and the colour vMMN ( $t(23) = -7.06$ ,  $p <$   
329  $.001$ ,  $BF_{10} = 6.3 \times 10^9$ ), suggesting that the vMMN is sensitive to features of the deviant  
330 stimulus.

### 331 *Feature-based Attention Decreases Neural Activity*

332 We assessed the main effect of feature-based attention by comparing ERPs to peripheral  
333 patterns that shared features with targets ('congruent') or distractors ('incongruent') in the

334 central detection task. Congruent peripheral patterns evoked a smaller positivity over posterior  
335 electrodes than incongruent patterns late in the epoch (188 – 305 ms,  $p = .004$ ; *Figure 3B,D*).  
336 This effect was matched by a polarity-reversed activity profile over frontal electrodes (191 –  
337 309 ms,  $p = .003$ ; *Figure 3A,D*).



338  
339 *Figure 3. Main effect of feature-based attention in Experiment 1. (A-B)* Congruent and incongruent  
340 ERPs are collapsed across prediction conditions separately for frontal electrodes (Fz, F1, F3, AFz,  
341 AF3, AF4; **A**) and occipital electrodes (Oz, O1, O2, POz, PO3, PO4; **B**). Shading indicates the within-  
342 subject standard error of the mean. Black bars along the x-axis denote significant differences at the  
343 displayed electrodes (cluster-corrected). **(C-D)** Headmaps show the effects of feature-based attention  
344 (congruent minus incongruent) during the indicated time windows. Asterisks and dots denote  
345 electrodes with larger, or smaller responses, respectively, in at least 33% of the averaged time points  
346 (cluster-corrected).

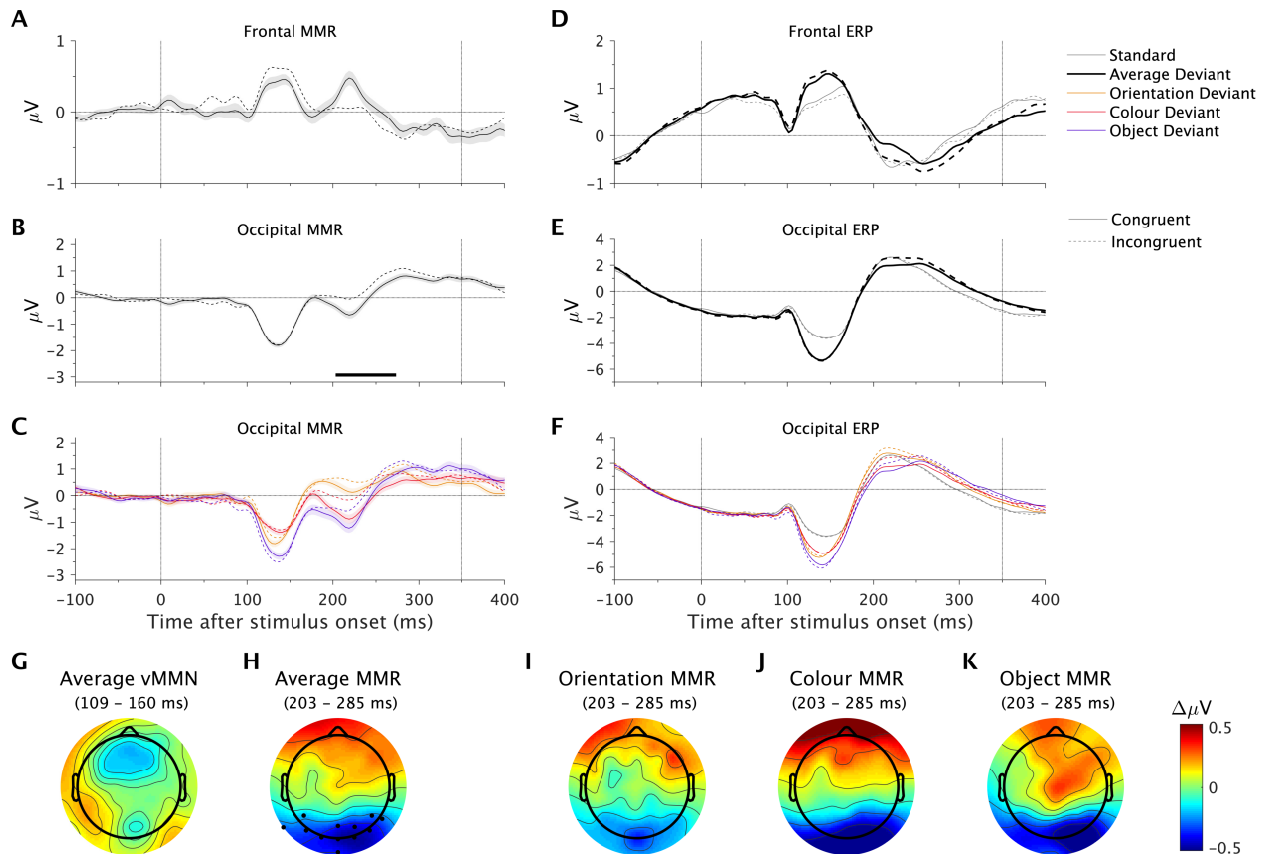
### 347 *The Effect of Feature-based Attention Depends on Surprise*

348 Next, we investigated the interaction between feature-based attention and prediction by  
349 subtracting the standard ERP from the deviant ERP (i.e., the mismatch response, collapsed  
350 across deviant conditions) and comparing difference waves between congruent and



351 incongruent conditions (*Figure 4*). Over posterior electrodes, the mismatch response was more  
352 negative for congruent stimuli than for incongruent stimuli late in the epoch (203 – 285 ms,  $p$   
353 = .041; *Figure 4B*). Inspection of individual condition ERPs (*Figure 4E*) revealed that the  
354 significant interaction was driven by a larger (negative) effect of feature-based attention on the  
355 neural response to deviants, relative to standards.

356 We followed up this finding by averaging spatiotemporal samples spanned by the significant  
357 effect, separately for each deviant condition. We then compared congruent and incongruent  
358 conditions with paired  $t$ -tests and Bayesian analyses (uniform prior with upper and lower  
359 bounds set to the average amplitude across all conditions). Feature-based attention decreased  
360 the mismatch response to all three deviant types (orientation: congruent =  $0.13 \pm 0.10 \mu\text{V}$ ,  
361 incongruent =  $0.42 \pm 0.13 \mu\text{V}$ ,  $t(23) = -2.16$ ,  $p = .041$ ,  $BF_{10} = 1.18$ ; colour: congruent =  $-0.45$   
362  $\pm 0.17 \mu\text{V}$ , incongruent =  $0.05 \pm 0.14 \mu\text{V}$ ,  $t(23) = -3.28$ ,  $p = .003$ ,  $BF_{10} = 1.41$ ; object: congruent  
363 =  $-0.53 \pm 0.19 \mu\text{V}$ , incongruent =  $-0.06 \pm 0.14 \mu\text{V}$ ,  $t(23) = -2.92$ ,  $p = .008$ ,  $BF_{10} = 1.28$ ; *Figure*  
364 *4C*). These findings suggest that feature-based attention modulates the effect of prediction on  
365 neural responses to stimuli at task-irrelevant locations, irrespective of the predicted feature (or  
366 combination of features).



367

368 *Figure 4. Interaction between feature-based attention and prediction in Experiment 1. (A-B) Average*  
 369 *mismatch response (MMR; average deviant minus standard) collapsed across frontal electrodes (Fz,*  
 370 *F1, F3, AFz, AF3, AF4; A) and occipital electrodes (Oz, O1, O2, POz, PO3, PO4; B). Solid lines*  
 371 *represent the congruent condition and dotted lines represent the incongruent condition. Shading*  
 372 *indicates the within-subject standard error of the mean. The black bar along the x-axis denotes*  
 373 *significant differences at the displayed electrodes (cluster-corrected). (C) Mismatch responses at*  
 374 *occipital electrodes for individual deviant conditions. (D-E) ERPs evoked by standards and deviants*  
 375 *(averaged across deviant types), shown separately for congruent (solid) and incongruent (dotted)*  
 376 *conditions. (F) ERPs for individual deviant conditions, shown separately for congruent (solid) and*  
 377 *incongruent (dotted) conditions. (G-H) Headmaps show the effect of feature-based attention*  
 378 *(congruent minus incongruent) on the average deviant mismatch response (average deviant minus*  
 379 *standard) during the early vMMN (G) and late interaction time windows (H). Dots denote electrodes*  
 380 *with significant differences in at least 33% of the averaged time points (cluster-corrected). (I-K) Effect*  
 381 *of feature-based attention (congruent minus incongruent) on the orientation mismatch response (I),*  
 382 *colour mismatch response (J) and object mismatch response (K) during the late interaction time*  
 383 *window. Note that cluster-based permutation tests were not conducted on these differences.*

384 *The Visual Mismatch Negativity is Not Modulated by Feature-based Attention*

385 Because previous literature has provided evidence for an effect of feature-based attention on  
 386 the vMMN (Czigler & Sulykos, 2010), we also used Bayes analyses to test for differences  
 387 between congruent and incongruent conditions during the (non-significant) vMMN time

388 period. Spatiotemporal samples spanning the common vMMN window (electrodes: Pz, P1, P2,  
389 P3, P4, POz, PO3, PO4, Oz, O1, O2, Iz; timepoints: 109 – 160 ms) were averaged to produce  
390 one amplitude value for each condition within participants. Congruent and incongruent  
391 conditions were compared within deviant conditions using paired-samples *t*-tests and Bayes  
392 analyses (uniform prior with upper and lower bounds set to the average amplitude across  
393 conditions). We found no difference between congruent and incongruent vMMNs for any  
394 deviant type (orientation: congruent =  $-1.09 \pm 0.13 \mu\text{V}$ , incongruent =  $-0.95 \pm 0.16 \mu\text{V}$ ,  $t(23) =$   
395  $-1.23$ ,  $p = .231$ ,  $BF_{10} = .26$ ; colour: congruent =  $-1.00 \pm 0.16 \mu\text{V}$ , incongruent =  $-0.96 \pm 0.14$   
396  $\mu\text{V}$ ,  $t(23) = -0.37$ ,  $p = .713$ ,  $BF_{10} = 0.13$ ; object: congruent =  $-1.55 \pm 0.19 \mu\text{V}$ , incongruent = -  
397  $1.73 \pm 0.15 \mu\text{V}$ ,  $t(23) = 1.29$ ,  $p = .208$ ,  $BF_{10} = 0.33$ ; *Figure 4C*).

398 Taken together, the results from Experiment 1 suggest that feature-based attention modulates  
399 the neural effect of prediction on neural responses to stimuli at task-irrelevant locations. This  
400 interaction emerged after (but not during) the vMMN time period for all deviant types, from  
401 approximately 200 ms after stimulus onset. We also found that the detection of high contrast  
402 targets at fixation was slower following peripheral patterns with target features, relative to  
403 those with distractor features, suggesting that feature-congruent peripheral patterns ‘captured’  
404 attention to their location (Folk et al., 1992). Because our principle question of interest  
405 pertained to the neural interaction between feature-based attention and prediction *outside* the  
406 current spatial focus of attention, we conducted a second study in which target contrast was  
407 individually titrated for each participant to increase the task difficulty and ensure that attention  
408 remained fixed on the central target stream.

409 **Experiment 2: Replication with individually thresholded manipulation of feature-based**  
410 **attention at fixation**

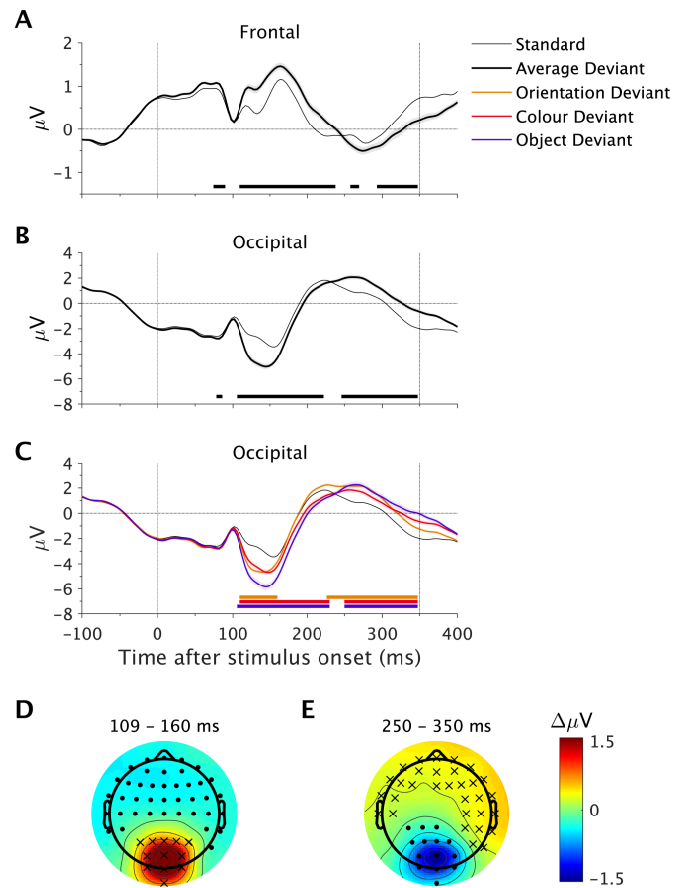
411 *Peripheral Patterns Do Not Modulate Behaviour in a Demanding Feature-based Attention*

412 *Task*

413 In contrast to Experiment 1, there was no significant effect of feature-congruence on  
414 normalised reaction times in Experiment 2 (congruent:  $391.79 \pm 11.12$  ms,  $-0.01 \pm 0.18$  z-  
415 normalised; incongruent:  $392.49 \pm 11.56$  ms,  $0.02 \pm 0.18$  z-normalised;  $F(1,23) = 1.00$ ,  $p =$   
416  $.329$ ,  $\eta_p^2 = .04$ ), suggesting that the top-down feature set modulates the effect of congruent  
417 patterns on target detection and that the more difficult task in Experiment 2 contained spatial  
418 attention to the central target stream. In line with Experiment 1, all other behavioural effects  
419 were non-significant. Thus, there was no significant effect of feature-congruence on  
420 normalised hit rates (congruent =  $74.90 \pm 2.05\%$ ,  $-0.02 \pm 0.16$  z-normalised, incongruent =  
421  $75.23 \pm 2.14\%$ ,  $0.01 \pm 0.17$  z-normalised,  $F(1,23) = 0.37$ ,  $p = .547$ ,  $\eta_p^2 = .02$ ). In addition, there  
422 was no significant effect of pattern prediction on normalised hit rates (predicted =  $74.25 \pm$   
423  $2.07\%$ ,  $-0.06 \pm 0.16$  z-normalised, surprising =  $75.88 \pm 2.13\%$ ,  $0.06 \pm 0.17$  z-normalised,  
424  $F(1,23) = 3.64$ ,  $p = .069$ ,  $\eta_p^2 = .14$ ) or on normalised reaction times (predicted =  $388.79 \pm 11.11$   
425 ms,  $-0.03 \pm 0.18$  z-normalised, surprising =  $395.49 \pm 11.58$  ms,  $0.04 \pm 0.18$  z-normalised,  
426  $F(1,23) = 2.46$ ,  $p = .130$ ,  $\eta_p^2 = .10$ ). Finally, there was no interaction between prediction and  
427 feature-congruence on either normalised hit rates ( $F(1,23) = 2.11$ ,  $p = .160$ ,  $\eta_p^2 = .08$ ) or  
428 normalised reaction times ( $F(1,23) = 1.50$ ,  $p = .233$ ,  $\eta_p^2 = .06$ ).

429 *The Neural Interaction Between Feature-based Attention and Prediction Replicates With a*  
430 *Demanding Feature-based Attention Set*

431 The neural effects observed in Experiment 2 (see *Figures 5-7*) were highly similar to those in  
432 Experiment 1 (see *Figures 2-4*). Prediction again modulated neural responses over posterior  
433 electrodes early (standards > deviants; from 78 ms,  $p < .001$ ) and late (standards < deviants;  
434 246 – 348 ms,  $p = .014$ ) in the epoch (*Figure 5B*), with opposite early (standards < deviants;  
435 74 - 238 ms,  $p < .001$ ) and late effects (standards > deviants; prior to 348 ms,  $p < .001$ ) over  
436 frontal electrodes (*Figure 5A*). Follow-up comparisons revealed similar effects of prediction  
437 on each deviant type (*Figure 5C*). Over posterior electrodes, standards evoked smaller early  
438 negativities than all deviant types (orientation deviants: 109 – 160 ms,  $p = .037$ ; colour  
439 deviants: 90 – 348 ms,  $p < .001$ ; object deviants; 78 – 348 ms,  $p < .001$ ) and smaller late  
440 positivities than all deviant types (orientation deviants: 227 – 348 ms,  $p = .001$ ; colour deviants;  
441 250 – 348 ms,  $p = .029$ ; object deviants: 250 – 348 ms,  $p = .022$ ). Over frontal electrodes,  
442 standards evoked smaller early positivities than all deviant types (orientation: 47 – 156 ms,  $p$   
443 = .0012; colour: 98 – 242 ms,  $p = .002$ ; object: 78 – 242 ms,  $p < .001$ ) and smaller late  
444 negativities than orientation deviants (234 – 348 ms,  $p < .001$ ). As in Experiment 1, the vMMN  
445 was sensitive to features of the deviant stimulus (*Figure 5C*), with object deviants evoking a  
446 significantly larger vMMN ( $-1.86 \pm 0.27 \mu\text{V}$ ) than orientation deviants ( $-1.10 \pm 0.16 \mu\text{V}$ ,  $t(23)$   
447 =  $-5.40$ ,  $p < .001$ ,  $BF_{10} = 288,942.02$ ) and colour deviants ( $-1.05 \pm 0.19 \mu\text{V}$ ,  $t(23) = -6.17$ ,  $p <$   
448  $.001$ ,  $BF_{10} = 22,207,026.78$ ). Again, there was no difference between orientation and colour  
449 vMMNs ( $t(23) = -0.42$ ,  $p = .679$ ,  $BF_{10} = 0.12$ ).



450

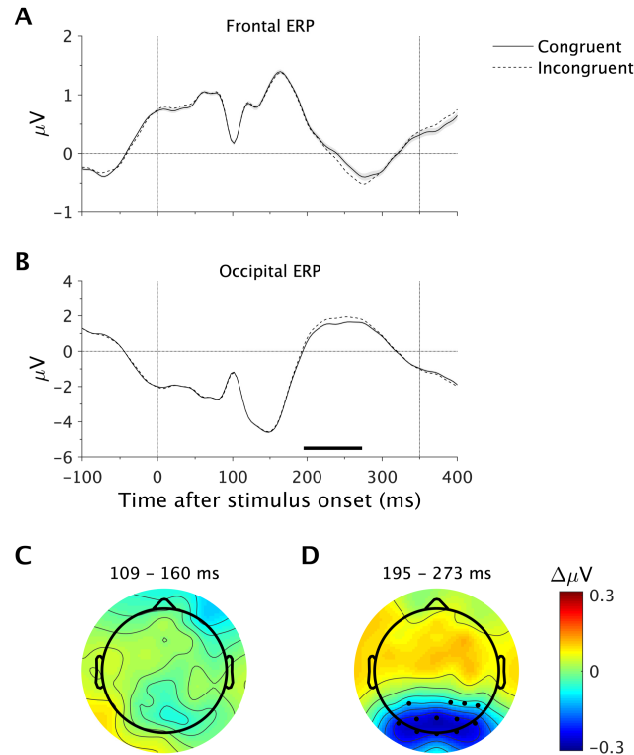
451 *Figure 5. Main effect of prediction in Experiment 2. (A-B) ERPs evoked by standards and deviants*  
 452 *(collapsed across deviant types) at frontal electrodes (Fz, F1, F3, AFz, AF3, AF4; A) and occipital*  
 453 *electrodes (Oz, O1, O2, POz, PO3, PO4; B). Shading indicates the within-subject standard error of the*  
 454 *mean, calculated relative to standards. Black bars along the x-axis denote significant timepoints at*  
 455 *the displayed electrodes (cluster-corrected). (C) ERPs evoked by standards and each of the three*  
 456 *deviant conditions. Shading indicates the within-subject standard error of the mean, calculated*  
 457 *separately for each deviant condition relative to standards. Yellow, red and purple bars along the x-*  
 458 *axis denote significant differences between standards and each corresponding deviant condition*  
 459 *(cluster-corrected). (D-E) Headmaps show the effect of prediction (standard minus average deviant)*  
 460 *during the indicated time windows. Asterisks and dots denote electrodes with larger, or smaller*  
 461 *responses, respectively, across at least 33% of the averaged time points (cluster-corrected).*

462 As in Experiment 1, congruent peripheral patterns evoked smaller positivities over posterior

463 electrodes than incongruent patterns late in the epoch (195 – 273 ms,  $p = .045$ ; *Figure 6B*).

464 However, the polarity-reversed frontal effect observed in Experiment 1 was not significant in

465 Experiment 2 (230 – 281 ms,  $p = .127$ ).



466

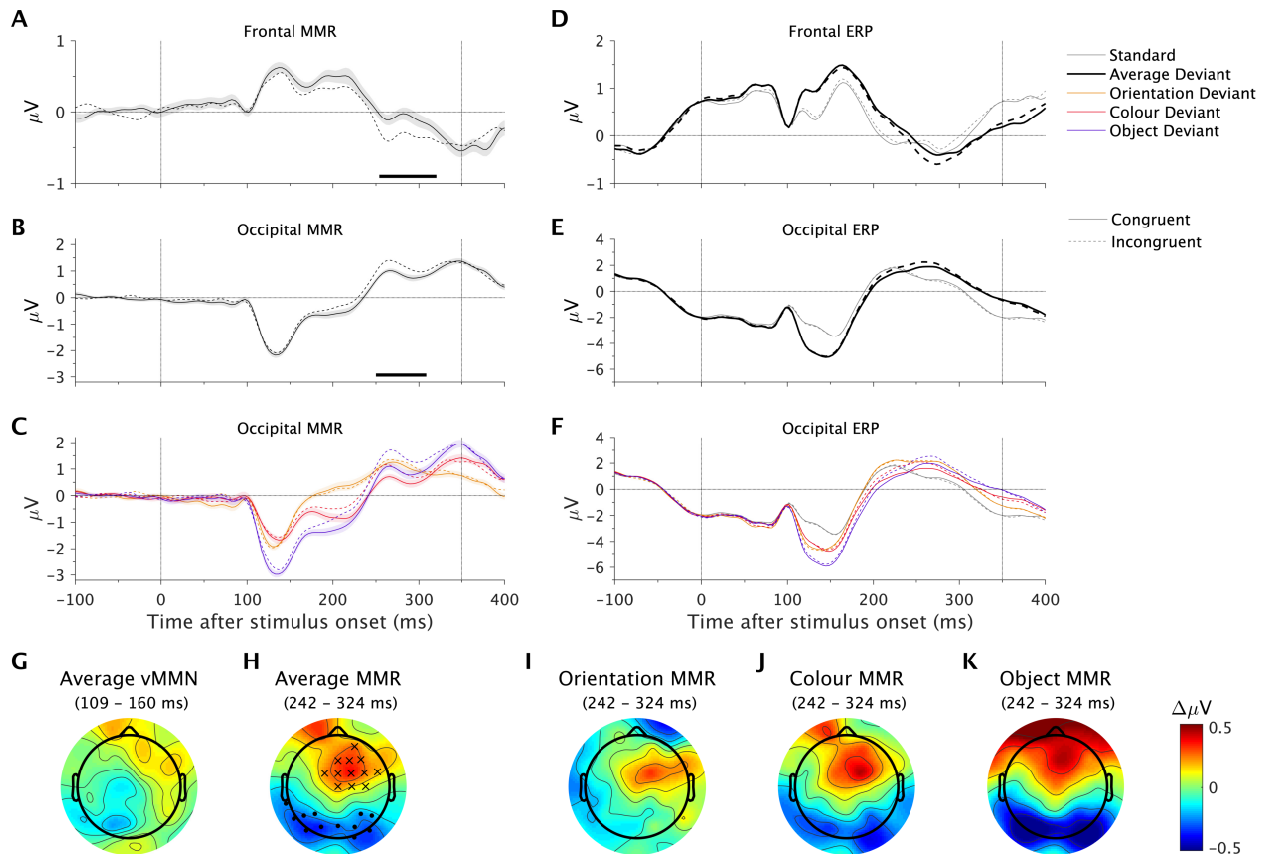
467 *Figure 6. Main effect of feature-based attention in Experiment 2. (A-B)* Congruent and incongruent  
468 ERPs are collapsed across prediction conditions separately for frontal electrodes (Fz, F1, F3, AFz,  
469 AF3, AF4; **A**) and occipital electrodes (Oz, O1, O2, POz, PO3, PO4; **B**). Shading indicates the within-  
470 subject standard error of the mean. The black bar along the x-axis denotes significant differences at  
471 the displayed electrodes (cluster-corrected). (**C-D**) Headmaps show the effects of feature-based  
472 attention (congruent minus incongruent) during the indicated time windows. Dots denote electrodes  
473 with smaller responses in at least 33% of the averaged time points (cluster-corrected).

474 Crucially, we replicated the significant interaction between feature-based attention and  
475 prediction observed in Experiment 1 (*Figure 7*). Congruent mismatch responses (deviants  
476 minus standards) were significantly smaller than incongruent mismatch responses over  
477 posterior electrodes late in the epoch (242 – 320 ms,  $p = .048$ ; *Figure 7B,H*). We also observed  
478 an additional polarity-reversed effect over frontal electrodes that was absent in Experiment 1  
479 (254 – 324 ms,  $p = .026$ ; *Figure 7A,H*). Follow-up analyses revealed that feature-based  
480 attention significantly decreased the mismatch response to colour deviants (congruent =  $-0.01$   
481  $\pm 0.17 \mu\text{V}$ , incongruent =  $0.31 \pm 0.13 \mu\text{V}$ ,  $t(23) = -3.52$ ,  $p = .002$ ,  $BF_{10} = 94.83$ , *Figure 7J*) and  
482 object deviants (congruent =  $0.10 \pm 0.21 \mu\text{V}$ , incongruent =  $0.61 \pm 0.14 \mu\text{V}$ ,  $t(23) = -3.66$ ,  $p =$

483 .001,  $BF_{10} = 51.42$ , *Figure 7K*) but only trended in the same direction for orientation deviants  
484 (congruent =  $0.40 \pm 0.11 \mu\text{V}$ , incongruent =  $0.60 \pm 0.10 \mu\text{V}$ ,  $t(23) = -1.96$ ,  $p = .062$ ,  $BF_{10} =$   
485  $2.35$ , *Figure 7I*). Again, we found no effect of feature-based attention on the vMMN evoked  
486 by any type of deviant (orientation: congruent =  $-1.09 \pm 0.19 \mu\text{V}$ , incongruent =  $-1.11 \pm 0.15$   
487  $\mu\text{V}$ ,  $t(23) = .18$ ,  $p = .857$ ,  $BF_{10} = .12$ ; colour: congruent =  $-1.08 \pm 0.19 \mu\text{V}$ , incongruent =  $-1.02$   
488  $\pm 0.22 \mu\text{V}$ ,  $t(23) = -0.42$ ,  $p = .676$ ,  $BF_{10} = .15$ ; object: congruent =  $-1.96 \pm 0.28 \mu\text{V}$ , incongruent  
489 =  $-1.76 \pm 0.27 \mu\text{V}$ ,  $t(23) = -1.67$ ,  $p = .11$ ,  $BF_{10} = 0.43$ ; *Figure 7C*).

490 Overall, the findings from Experiment 2 replicate those from Experiment 1 to show that  
491 feature-based attention and prediction interact in their modulation of neural responses to stimuli  
492 at task-irrelevant locations. Similar to Experiment 1, this interaction emerged after (but not  
493 during) the vMMN time period for all deviant types, from approximately 240 ms after stimulus  
494 onset.





495

496 *Figure 7. Interaction between feature-based attention and prediction in Experiment 2. (A-B) Average*  
 497 *mismatch response (MMR; average deviant minus standard) collapsed across frontal electrodes (Fz,*  
 498 *F1, F3, AFz, AF3, AF4; A) and occipital electrodes (Oz, O1, O2, POz, PO3, PO4; B). Solid lines*  
 499 *represent the congruent condition and dotted lines represent the incongruent condition. Shading*  
 500 *indicates the within-subject standard error of the mean. The black bar along the x-axis denotes*  
 501 *significant differences at the displayed electrodes (cluster-corrected). (C) Mismatch responses at*  
 502 *occipital electrodes for individual deviant conditions. (D-E) ERPs evoked by standards and deviants*  
 503 *(averaged across deviant types), shown separately for congruent (solid) and incongruent (dotted)*  
 504 *conditions. (F) ERPs for individual deviant conditions, shown separately for congruent (solid) and*  
 505 *incongruent (dotted) conditions. (G-H) Headmaps show the effect of feature-based attention*  
 506 *(congruent minus incongruent) on the average deviant mismatch response (average deviant minus*  
 507 *standard) during the early vMMN (G) and late interaction time windows (H). Asterisks and dots denote*  
 508 *electrodes with larger, or smaller responses, respectively, in at least 33% of the averaged time points*  
 509 *(cluster-corrected). (I-K) Effect of feature-based attention (congruent minus incongruent) on the*  
 510 *orientation mismatch response (I), colour mismatch response (J) and object mismatch response (K)*  
 511 *during the late interaction time window. Note that cluster-based permutation tests were not conducted*  
 512 *on these differences.*

513

## Discussion

514 Here we investigated whether prediction interacts with feature-based attention outside the  
 515 spatial focus of attention. To achieve this, we measured neural responses to surprising and  
 516 predicted stimuli – deviants and standards, respectively – presented at task-irrelevant locations.

517 Task-irrelevant peripheral patterns shared features with either the targets (congruent) or  
518 distractors (incongruent) in a central search task. Across two experiments, we replicated the  
519 finding that feature-based attention decreased neural responses to surprising but not predicted  
520 task-irrelevant stimuli in the periphery of vision. This finding suggests that the global neural  
521 mechanisms of feature-based attention and prediction are interdependent, and supports the  
522 theory that attention increases the gain of prediction errors (Feldman & Friston, 2010).

523 Consistent with previous literature, prediction modulated early and late neural responses to  
524 stimuli in both experiments (*Figures 2 & 5*). Early responses (approximately 100 to 160 ms)  
525 over posterior electrodes were more negative for surprising stimuli than predicted stimuli,  
526 consistent with the commonly reported visual mismatch negativity (for a review, see Stefanics  
527 et al., 2014). This finding is broadly consistent with the theory that top-down prediction signals  
528 silence matching bottom-up sensory signals and leave only the remaining prediction error to  
529 propagate forward (Friston, 2005, 2009; Rao & Ballard, 1999). Prediction also reduced the  
530 later positive P3 component (from approximately 250 ms), consistent with the theory that this  
531 component reflects involuntary orienting to novel stimuli (Friedman, Cycowicz, & Gaeta,  
532 2001; Polich, 2007).

533 We also found that stimuli deviating in two feature dimensions (i.e., object deviants) evoked  
534 larger early negativities than stimuli deviating in only one feature dimension (i.e., orientation  
535 or colour deviants; *Figures 2 & 5*). This finding contradicts a previous study that found visual  
536 features elicit non-additive mismatch-related brain activity (Sulykos & Czigler, 2011), and  
537 suggests instead that the vMMN is sensitive to the extent of deviation across multiple feature  
538 dimensions. Importantly, object deviants in Sulykos & Czigler (2011) deviated in spatial  
539 frequency and orientation, whereas object deviants in the present study deviated in colour and

540 orientation. Thus, future studies should investigate the extent to which mismatch additivity in  
541 the visual domain depends on the specific features involved.

542 We found that feature-based attention reduced neural responses to task-irrelevant peripheral  
543 patterns from approximately 200 ms after stimulus onset (*Figures 3 & 6*), consistent with the  
544 commonly reported ‘selection negativity’ (Gledhill et al., 2015). This effect replicated with  
545 low contrast stimuli that likely necessitated a tight focus of spatial attention (Experiment 2),  
546 contradicting the finding that feature-specific modulation of the selection negativity is  
547 contingent on spatial attention (Anllo-Vento & Hillyard, 1996; Hillyard & Münte, 1984) and  
548 suggesting instead that late effects of feature-based attention are globally effective (Gledhill et  
549 al., 2015). Interestingly, we found no difference between neural responses to congruent and  
550 incongruent stimuli earlier in the epoch (*Figures 3 & 6*), in contrast to a previous study that  
551 reported early effects of feature-based attention on neural responses to stimuli at task-irrelevant  
552 locations (beginning within 100 ms of stimulus onset; Zhang & Luck, 2009). A critical  
553 difference between Zhang & Luck (2009) and the present study is that Zhang & Luck (2009)  
554 had participants search for targets with specific feature conjunctions (luminance and colour),  
555 whereas targets in our study were defined by only a single feature (colour or orientation). Thus,  
556 it is possible that early effects of feature-based attention depend on the complexity of the  
557 attentional set. This interpretation is consistent with a recent study in which we found that  
558 neural responses to high-frequency flickering stimuli outside a search array (12.5 or 16.7 Hz,  
559 corresponding to an 80 or 60 ms cycle) are enhanced by feature-based attention during  
560 conjunction but not unique-feature search (Painter, Dux, Travis, & Mattingley, 2014).

561 Crucially, we found an interaction between feature-based attention and prediction in each of  
562 the two experiments. Congruent stimuli evoked smaller posterior mismatch responses than

563 incongruent stimuli between approximately 200 and 300 ms after stimulus onset. Inspection of  
564 the ERPs revealed that the effect of feature-based attention on neural responses was larger for  
565 deviants than it was for standards. This pattern of results is consistent with our recent finding  
566 that attention enhances the processing of mismatch information from approximately 200 ms  
567 post-stimulus (Smout et al., 2019) and broadly supports the theory that attention enhances the  
568 gain of prediction errors (Feldman & Friston, 2010). Neural responses to surprising stimuli  
569 (deviants) are theorised to be modulated by attention because they contain prediction errors,  
570 whereas neural responses to predicted stimuli (standards) are less affected because they contain  
571 relatively few prediction errors. The present study extends this theory to suggest that feature-  
572 specific attentional modulation of prediction errors occurs even when the surprising stimuli are  
573 task-irrelevant and presented outside the spatial focus of attention.

574 Interestingly, we found that feature-based attention had no effect on the earlier vMMN evoked  
575 by deviants (109 – 160 ms). This pattern of findings contradicts a previous study that found the  
576 vMMN evoked by peripheral stimuli was smaller (more positive) when participants searched  
577 for a change in the deviating feature at fixation, relative to a different feature (Czigler &  
578 Sulykos, 2010). A subtle difference between the paradigms is that participants in Czigler and  
579 Sulykos (2010) searched for a feature ‘change’ at fixation (e.g., a change in the target object  
580 colour), whereas participants in our study searched for specific object onsets. Thus, it remains  
581 possible that subtle differences in the configuration of the attentional set can influence the  
582 timing and direction of the interaction between feature-based attention and prediction.

583 We manipulated target and distractor salience across the two experiments in order to  
584 investigate whether the strength of the top-down feature set modulates the neural interaction  
585 between prediction and feature-based attention. Although the pattern of neural effects did not

586 differ between the two experiments, we observed slightly different behavioural effects as a  
587 function of task difficulty. Responses to highly salient targets (Experiment 1) that appeared  
588 immediately after a congruent pattern were slower than those to targets that appeared after an  
589 incongruent pattern. In contrast, there was no such effect of feature-congruence on responses  
590 to less salient targets (Experiment 2). These findings are broadly consistent with contingent  
591 capture theory (Folk et al., 1992), which proposes that distracting stimuli within the spatial  
592 focus of attention capture attention when they are congruent with the observers' current  
593 attentional set. Since targets were easily detected in Experiment 1, it seems likely that some  
594 amount of spatial attention 'leaked' to the peripheral stimuli, facilitating contingent capture. In  
595 contrast, the higher task difficulty of Experiment 2 likely necessitated a tighter focus of  
596 attention to the central stimuli, thus prohibiting a contingent capture effect.

597 We did not observe an effect of predictability of peripheral patterns on target detection, or an  
598 interaction between pattern prediction and feature-congruence, in either experiment. This is  
599 consistent with a previous study that failed to find any effect of pattern prediction on response  
600 times to a central feature change target, nor an interaction with task set, at the level of single  
601 trials (though note that this study did report sustained block-wise effects on behaviour; Czigler  
602 & Sulykos, 2010). These findings suggest that the neural bias toward feature-congruent and  
603 surprising stimuli at task-irrelevant locations, observed in the present study, does not interfere  
604 with the concurrent processing of targets at task-relevant locations.

605 The present study contributes to a burgeoning literature on the relationship between prediction  
606 and attention. Whereas some studies have found an interaction between prediction and  
607 attention (Auksztulewicz & Friston, 2015; Jiang et al., 2013; Kok, Rahnev, et al., 2012;  
608 Marzecová et al., 2017; Smout et al., 2019), many others have reported only independent main

609 effects (e.g. Garrido, Rowe, Halász, & Mattingley, 2017; Hsu, Hämäläinen, & Waszak, 2014;  
610 Kok, Jehee, & de Lange, 2012). We note that investigations to date have employed a wide  
611 variety of attention manipulations (e.g., feature-based, spatial, temporal) and prediction  
612 manipulations (e.g., first-order, rule-based) across different sensory modalities (e.g., visual,  
613 auditory). Thus, the equivocal pattern of findings to date may stem from distinct relationships  
614 between different subprocesses of attention and prediction across the various modalities. In  
615 particular, previous studies that found an interaction between visual attention and prediction  
616 presented stimuli at attended locations (Jiang et al., 2013; Kok, Rahnev, et al., 2012; Marzecová  
617 et al., 2017; Smout et al., 2019) or used paradigms that did not require focussed attention to  
618 complete the task (Czigler & Sulykos, 2010), leaving open the possibility that spatial attention  
619 is necessary in the interaction with prediction. The present study extends this literature by  
620 demonstrating that visual predictions interact with feature-based attention to modulate neural  
621 responses to stimuli outside the spatial focus of attention. The nature of this interaction is  
622 consistent with the theory that attention optimises the expected precision of predictions by  
623 modulating the gain of prediction errors (Feldman & Friston, 2010). Future research should  
624 continue to parse ‘attention’ and ‘prediction’ into more precise taxonomies that reflect specific  
625 mechanisms in the brain and investigate potential interactions between each of these  
626 subcomponents. This work could illuminate the extent to which predictive coding theory might  
627 be considered a ‘unified theory of the brain’ (Friston, 2010).

## 628 **References**

- 629 Anllo-Vento, L., & Hillyard, S. A. (1996). Selective attention to the color and direction of  
630 moving stimuli: Electrophysiological correlates of hierarchical feature selection.  
631 *Perception & Psychophysics*, 58(2), 191–206. <https://doi.org/10.3758/BF03211875>  
632 Anstis, S. M., & Cavanagh, P. (1983). A minimum motion technique for judging

- 633           equiluminance. *Colour Vision: Physiology and Psychophysics*, 155–166.
- 634   Auksztulewicz, R., & Friston, K. (2015). Attentional Enhancement of Auditory Mismatch  
635           Responses: a DCM/MEG Study. *Cerebral Cortex*, 1–11.  
636           <https://doi.org/10.1093/cercor/bhu323>
- 637   Carrasco, M. (2011). Visual attention: The past 25 years. *Vision Research*, 51(13), 1484–  
638           1525.
- 639   Chaumon, M., Bishop, D. V. M., & Busch, N. A. (2015). A practical guide to the selection of  
640           independent components of the electroencephalogram for artifact correction. *Journal of*  
641           *Neuroscience Methods*, 250, 47–63. <https://doi.org/10.1016/j.jneumeth.2015.02.025>
- 642   Czigler, I., & Sulykos, I. (2010). Visual mismatch negativity to irrelevant changes is sensitive  
643           to task-relevant changes. *Neuropsychologia*, 48(5), 1277–1282.  
644           <https://doi.org/10.1016/j.neuropsychologia.2009.12.029>
- 645   Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-  
646           trial EEG dynamics including independent component analysis. *Journal of Neuroscience*  
647           *Methods*, 134, 9–21.
- 648   Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual*  
649           *Review of Neuroscience*, 18(1), 193–222.  
650           <https://doi.org/10.1146/annurev.ne.18.030195.001205>
- 651   Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in*  
652           *Psychology*, 5(July), 1–17. <https://doi.org/10.3389/fpsyg.2014.00781>
- 653   Feldman, H., & Friston, K. J. (2010). Attention, uncertainty, and free-energy. *Frontiers in*  
654           *Human Neuroscience*, 4(December), 215. <https://doi.org/10.3389/fnhum.2010.00215>
- 655   Folk, C. L., Remington, R. W., & Johnston, J. C. (1992). Involuntary Covert Orienting Is  
656           Contingent on Attentional Control Settings. *Journal of Experimental Psychology:*  
657           *Human Perception and Performance*, 18(4), 1030–1044. [https://doi.org/10.1037/0096-](https://doi.org/10.1037/0096-1523.18.4.1030)  
658           1523.18.4.1030
- 659   Friedman, D., Cycowicz, Y. M., & Gaeta, H. (2001). The novelty P3: An event-related brain  
660           potential (ERP) sign of the brain's evaluation of novelty. *Neuroscience and*  
661           *Biobehavioral Reviews*, 25(4), 355–373. [https://doi.org/10.1016/S0149-7634\(01\)00019-](https://doi.org/10.1016/S0149-7634(01)00019-7)  
662           7

- 663 Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal*  
664 *Society B: Biological Sciences*, 360(1456), 815–836.
- 665 Friston, K. (2009). The free-energy principle: a rough guide to the brain? *Trends in Cognitive*  
666 *Sciences*, 13(7), 293–301. <https://doi.org/10.1016/j.tics.2009.04.005>
- 667 Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature Reviews.*  
668 *Neuroscience*, 11(2), 127–138. <https://doi.org/10.1038/nrn2787>
- 669 García-Pérez, M. A. (1998). Forced-choice staircases with fixed step sizes: asymptotic and  
670 small-sample properties. *Vision Research*, 38(12), 1861–1881.
- 671 Garrido, M. I., Rowe, E. G., Halász, V., & Mattingley, J. B. (2017). Bayesian Mapping  
672 Reveals That Attention Boosts Neural Responses to Predicted and Unpredicted Stimuli,  
673 1–12. <https://doi.org/10.1093/cercor/bhx087>
- 674 Gledhill, D., Grimsen, C., Fahle, M., & Wegener, D. (2015). Human feature-based attention  
675 consists of two distinct spatiotemporal processes. *Journal of Vision*, 15(8), 1–17.  
676 <https://doi.org/10.1167/15.8.8.doi>
- 677 Hillyard, S., & Münte, T. (1984). Selective attention to color and location: an analysis with  
678 event-related brain potentials. *Perception & Psychophysics*, 36(2), 185–198.  
679 <https://doi.org/10.3758/BF03202679>
- 680 Hohwy, J. (2012). Attention and conscious perception in the hypothesis testing brain.  
681 *Frontiers in Psychology*, 3(APR), 1–14. <https://doi.org/10.3389/fpsyg.2012.00096>
- 682 Hsu, Y., Hämäläinen, J. A., & Waszak, F. (2014). Both attention and prediction are necessary  
683 for adaptive neuronal tuning in sensory processing. *Frontiers in Human Neuroscience*,  
684 8(March), 152. <https://doi.org/10.3389/fnhum.2014.00152>
- 685 Jiang, J., Summerfield, C., & Egner, T. (2013). Attention sharpens the distinction between  
686 expected and unexpected percepts in the visual brain. *The Journal of Neuroscience : The*  
687 *Official Journal of the Society for Neuroscience*, 33(47), 18438–18447.  
688 <https://doi.org/10.1523/JNEUROSCI.3308-13.2013>
- 689 Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007). What's  
690 new in Psychtoolbox-3. *Perception*, 36(14), 1.
- 691 Kok, P., Jehee, J. F. M., & de Lange, F. P. (2012). Less is more: Expectation sharpens  
692 representations in the primary visual cortex. *Neuron*, 75, 265–270.



- 693 <https://doi.org/10.1016/j.neuron.2012.04.034>
- 694 Kok, P., Rahnev, D., Jehee, J. F. M., Lau, H. C., & De Lange, F. P. (2012). Attention reverses  
695 the effect of prediction in silencing sensory signals. *Cerebral Cortex*, 22(9), 2197–2206.  
696 <https://doi.org/10.1093/cercor/bhr310>
- 697 Martinez-Trujillo, J. C., & Treue, S. (2004). Feature-Based Attention Increases the  
698 Selectivity of Population Responses in Primate Visual Cortex. *Current Biology*, 14(9),  
699 744–751. <https://doi.org/10.1016/j.cub.2004.04.028>
- 700 Marzecová, A., Widmann, A., SanMiguel, I., Kotz, S. A., & Schröger, E. (2017).  
701 Interrelation of attention and prediction in visual processing: Effects of task-relevance  
702 and stimulus probability. *Biological Psychology*, 125, 76–90.  
703 <https://doi.org/10.1016/j.biopsycho.2017.02.009>
- 704 Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J. M. (2011). FieldTrip: Open source  
705 software for advanced analysis of MEG, EEG, and invasive electrophysiological data.  
706 *Computational Intelligence and Neuroscience*, 2011, 1–9.  
707 <https://doi.org/10.1155/2011/156869>
- 708 Painter, D. R., Dux, P. E., Travis, S. L., & Mattingley, J. B. (2014). Neural responses to  
709 target features outside a search array are enhanced during conjunction but not unique-  
710 feature search. *Journal of Neuroscience*, 34(9), 3390–3401.  
711 <https://doi.org/10.1523/JNEUROSCI.3630-13.2014>
- 712 Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clinical*  
713 *Neurophysiology*, 118(10), 2128–2148. <https://doi.org/10.1016/j.clinph.2007.04.019>
- 714 Posner, M. I. (1994). Attention: the mechanisms of consciousness. *Proceedings of the*  
715 *National Academy of Sciences of the United States of America*, 91(16), 7398–7403.
- 716 Rao, R. P. N., & Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional  
717 interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, 2(1),  
718 79–87. <https://doi.org/10.1038/4580>
- 719 Saenz, M., Buracas, G. T., & Boynton, G. M. (2002). Global effects of feature-based  
720 attention in human visual cortex. *Nature Neuroscience*, 5(7), 631–632.
- 721 Serences, J. T., & Boynton, G. M. (2007). Feature-Based Attentional Modulations in the  
722 Absence of Direct Visual Stimulation. *Neuron*, 55(2), 301–312.

- 723 <https://doi.org/10.1016/j.neuron.2007.06.015>
- 724 Smout, C. A., Tang, M. F., Garrido, M. I., & Mattingley, J. B. (2019). Attention Promotes the  
725 Neural Encoding of Prediction Errors. *PLOS Biology*, 522185.  
726 <https://doi.org/10.1101/522185>
- 727 Stefanics, G., Kremláček, J., & Czigler, I. (2014). Visual mismatch negativity: a predictive  
728 coding view. *Frontiers in Human Neuroscience*, 8(September), 1–19.  
729 <https://doi.org/10.3389/fnhum.2014.00666>
- 730 Sulykos, I., & Czigler, I. (2011). One plus one is less than two: Visual features elicit non-  
731 additive mismatch-related brain activity. *Brain Research*, 1398, 64–71.  
732 <https://doi.org/10.1016/j.brainres.2011.05.009>
- 733 Zhang, W., & Luck, S. J. (2009). Feature-based attention modulates feedforward visual  
734 processing. *Nature Neuroscience*, 12(1), 24–25. <https://doi.org/10.1038/nn.2223>  
735